73 101 A

88 367712

Abnormality Detection and Surveillance System

1   **Field of the Invention**

2   This invention generally relates to surveillance systems, and more

3   particularly, to trainable surveillance systems which detect and respond to specific

4   abnormal video and audio input signals.

5

6   **Background of the Invention**

7   Today's surveillance systems vary in complexity, efficiency and accuracy.

8   Earlier surveillance systems use several closed circuit cameras, each connected to a

9   devoted monitor. This type of system works sufficiently well for low-coverage sites,

10   i.e., areas requiring up to perhaps six cameras. In such a system, a single person could

11   scan the six monitors, in "real" time, and effectively monitor the entire (albeit small)

12   protected area, offering a relatively high level of readiness to respond to an abnormal act

13   or situation observed within the protected area. In this simplest of surveillance systems,

14   it is left to the discretion of security personnel to determine, first, if there is any

15   abnormal event in progress within the protected area, second, the level of concern placed

16   on that particular event, and third, what actions should be taken in response to the

17   particular event. The reliability of the entire system depends on the alertness and

18   efficiency of the worker observing the monitors.

19   Many surveillance systems, however, require the use of a greater number

20   of cameras (e.g., more than six) to police a larger area, such as at least every room

21   located within a large museum. To adequately ensure reliable and complete surveillance

1   within the protected area, either more personnel must be employed to constantly watch

2   the additionally required monitors (one per camera), or fewer monitors may be used on

3   a simple rotation schedule wherein one monitor sequentially displays the output images

4   of several cameras, displaying the images of each camera for perhaps a few seconds.

5   In another prior art surveillance system (referred to as the "QUAD" system), four

6   cameras are connected to a single monitor whose screen continuously and simultaneously

7   displays the four different images. In a "quaded quad" prior art surveillance system,

8   sixteen cameras are linked to a single monitor whose screen now displays, continuously

9   and simultaneously all sixteen different images. These improvements allow fewer

10   personnel to adequately supervise the monitors to cover the larger protected area.

11   These improvements, however, still require the constant attention of at

12   least one person. The above described multiple-image/single screen systems suffered

13   from poor resolution and complex viewing. The reliability of the entire system is still

14   dependent on the alertness and efficiency of the security personnel watching the monitors.

15   The personnel watching the monitors are still burdened with identifying an abnormal act

16   or condition shown on one of the monitors, determining which camera, and which

17   corresponding zone of the protected area is recording the abnormal event, determining

18   the level of concern placed on the particular event, and finally, determining the

19   appropriate actions that must be taken to respond to the particular event.

20   Eventually, it was recognized that human personnel could not reliably

21   monitor the "real-time" images from one or several cameras for long "watch" periods

22   of time. It is natural for any person to become bored while performing a monotonous

1 task, such as staring at one or several monitors continuously, waiting for something

2 unusual or abnormal to occur, something which may never occur.

3 As discussed above, it is the human link which lowers the overall

4 reliability of the entire surveillance system. U.S. Patent 4,737,847 issued to Araki et al.

5 discloses an improved abnormality surveillance system wherein motion sensors are

6 positioned within a protected area to first determine the presence of an object of interest,

7 such as an intruder. In the system disclosed by U.S. Patent 4,737,847, zones having

8 prescribed "warning levels" are defined within the protected area. Depending on which

9 of these zones an object or person is detected in, moves to, and the length of time the

10 detected object or person remains in a particular zone determines whether the object or

11 person entering the zone should be considered an abnormal event or a threat.

12 The surveillance system disclosed in U.S. Patent 4,737,847 does remove

13 some of the monitoring responsibility otherwise placed on human personnel; however,

14 such a system can only determine an intruder's "intent" by his presence relative to

15 particular zones. The actual movements and sounds of the intruder are not measured or

16 observed. A skilled criminal could easily determine the warning levels of obvious zones

17 within a protected area and act accordingly; spending little time in zones having a high

18 warning level, for example.

19 It is therefore an object of the present invention to provide a surveillance

20 system which overcomes the problems of the prior art.

1   It is another object of the invention to provide such a surveillance system

2   wherein a potentially abnormal event is determined by a computer prior to summoning

3   a human supervisor.

4   It is another object of the invention to provide a surveillance system which

5   compares specific measured movements of a particular person or persons with a

6   trainable, predetermined set of "typical" movements to determine the level and type of

7   a criminal or mischievous event.

8   It is another object of this invention to provide a surveillance system which

9   transmits the data from various sensors to a location where it can be recorded for

10   evidentiary purposes.   It is another object of this invention to provide such

11   a surveillance system which is operational day and night.

12   It is another object of this invention to provide a surveillance system which

13   can cull out real-time events which indicate criminal intent using a weapon, by resolving

14   the low temperature of the weapon relative to the higher body temperature and by

15   recognizing the stances taken by the person with the weapon.

16   It is yet another object of this invention to provide a surveillance system

17   which eliminates or reduces the number of TV monitors and guards presently required

18   to identify abnormal events, as this system will perform this function in near real time.

19

20   **Incorporated by Reference**

21   The content of the following references is hereby incorporated by

22   reference.

1   1.   Motz L. and L. Bergstein "Zoom Lens Systems", Journal of

2   Optical Society of America, 3 papers in Vol. 52, 1992.

3   2.   D.G. Aviv, "Sensor Software Assessment of Advanced Earth

4   Resources Satellite Systems", ARC Inc. Report #70-80-A, pp. 2-107 through 2-119;

5   NASA contract NAS-1-16366.

6   3.   Shio, A. and J. Sklansky "Segmentation of People in Motion",

7   Proc. of IEEE Workshop on Visual Motion, Princeton, NJ, October 1991.

8   4.   Agarwal, R. and J Sklansky "Estimating Optical Flow from

9   Clustered Trajectory Velocity Time".

10   5.   Suzuki, S. and J Sklansky "Extracting Non-Rigid Moving Objects

11   by Temporal Edges", IEEE, 1992, Transactions of Pattern Recognition.

12   6.   Rabiner, L. and Biing-Hwang Juang "Fundamental of Speech

13   Recognition", Pub. Prentice Hall, 1993, (p.434-495).

14   7.   Weibel, A. and Kai-Fu Lee Eds. "Readings in Speech

15   Recognition", Pub. Morgan Kaaufman, 1990 (p.267-296).

16   8.   Rabiner, L. "Application of Voice Processing to

17   Telecommunication", Proc. IEEE, Vol. 82, No.2, February, 1994.

18

19   **Summary of the Invention**

20   A preferred embodiment of the herein disclosed invention involves a

21   surveillance system having at least one primary video camera for translating real images

22   of a zone into electronic video signals at a first level of resolution and means for

5

1   sampling movements within the zone from the video camera output. These elements are

2   combined with means for electronically comparing the sampled movements with known

3   characteristics of movements which are indicative of individuals engaged in criminal

4   activity and means for determining the level of such criminal activity. Associated

5   therewith are means for activating at least one secondary sensor and associated recording

6   device having a second higher level of resolution, said activating means being in response

7   to determining a predetermined level of criminal activity.

8

Brief Description of the Drawings

10   Figure 1 is a schematic block diagram of the video, analysis, control,

11   alarm and recording subsystems of an embodiment of this invention;

12   Figure 2 depicts an array of frames illustrating how the process of

13   segmentation and location of the different objects (people) in an embodiment of the

14   invention and identifies interaction between the objects (people);

15   Figure 3 depicts an array of frames illustrating a "two on one" interaction,

16   wherein two objects, (e.g., two persons) accost a third object (person); and

17   Figure 4 is a schematic block diagram of a conventional word recognition

18   system which may be employed in the invention.

19

Detailed Description of the Preferred Embodiments

21   Referring to Fig. 1, the picture input means 10, may be any conventional

22   electronic picture pickup device operational within the infrared or visual spectrum (or

both) including a vidicon and a CCD/TV camera of moderate resolution, e.g., a camera about 1 1/2 inches in length and about 1 inch in diameter, weighing about 3 ounces, including for particular deployment a zoom lens attachment. This device is intended to operate continuously and translate the field of view ("real") images within a first observation area into conventional video electronic signals.

Alternatively, a high rate camera/recorder, up to 300 frames/sec (similar to those made by NAC Visual Systems of Woodland Hills, CA, SONY and others) may be used as the picture input means 10. This would enable the detection of even the very rapid movement of body parts that are indicative of criminal intent, and their recording, as hereinbelow described. The more commonly used camera operates at 30 frames per second and cannot capture such quick body movement with sufficient resolution.

Picture input means 10, instead of operating continuously, may be activated by an "alert" signal from the processor of the low resolution camera or from the audio/word recognition processor when sensing a suspicious event.

Picture input means 10 contains a preprocessor which normalizes a wide range of illumination levels, especially for outside observation. The preprocessor emulates a vertebrate's retina, which has a an efficient and accurate normalization process. One such preprocessor (VLSI retina chip) is fabricated by the Carver Meade Laboratory of the California Institute of Technology in Pasadena, California. Use of this particular preprocessor chip will increase the automated vision capability of this invention whenever variation of light intensity and light reflection may otherwise weaken the picture resolution.

The signals from the picture input means 10 are converted into digitized signals and then sent to the picture processing means 12. The processor means controlling each group of cameras will be governed by an artificial intelligence system, based on dynamic pattern recognition principles, as further described below. Picture processing means 12 includes an image raster analyzer which effectively segments each image to isolate each pair of people. The image raster analyzer subsystem of picture processing means 12 segments each sampled image to identify and isolate each pair of objects (or people), and each "two on one" group of three people separately.

The "two on one" grouping represents a common mugging situation in which two individuals approach a victim, one from in front of the victim and the other from behind. The forward mugger tells the potential victim that if he does not give up his money, (or watch, ring, etc.) the second mugger will shoot him, stab or otherwise harm him. The group of three people will thus be considered a potential crime in progress and will therefore be segmented and analyzed in picture processing means.

With respect to a zoom lens system useful as an element in the picture input means 10, the essentials of the zoom lens subsystem are described in three papers written by L. Motz and L. Bergstein, in an article titled "Zoom Lens Systems" in the Journal of Optical Society of America, Vol. 52, April, 1992. This article is hereby incorporated by reference.

The essence of the zoom system is to vary the focal length such that an object being observed will be focused and magnified at its image plane. In an automatic version of the zoom system, once an object is in the camera's field-of-view (FOV), the

1 lens moves to focus the object onto the camera's image plane. An error signal which is

2 used to correct the focus by the image planes is generated by a CCD array into two

3 halves and measuring the difference, segmenting in each until the object is at the center.

4 Dividing the CCD array into more than two segments, say four quadrants, is a way to

5 achieve automatic centering, as is the case with mono-pulse radar. Regardless of the

6 number of segments, the error signal is used to generate the desired tracking of the

7 object.

8 In a wide field-of-view (WFOV) operation, there may be more than one

9 object, thus special attention is given to the design of the zoom system and its associated

10 software and firmware control. Assuming three objects, as is the "two on one" potential

11 mugging threat described above, and that the three persons are all in one plane, one can

12 program a shifting from one object to the next, from one face to another face, in a

13 prescribed sequential order. Moreover, as the objects move within the WFOV they will

14 be automatically tracked in azimuth and elevation. In principle, the zoom would focus

15 on the nearest object, assuming that the amount of light on each object is the same so

16 that the prescribed sequence starting from the closes object will proceed to the remaining

17 objects from, for example, right to left.

18 However, when the three objects are located in different planes, but still

19 within the camera's WFOV, the zoom, with input from the segmentation subsystem of

20 the picture analysis means 12 will focus on the object closest to the right hand side of the

21 image plane, and then proceed to move the focus to the left, focusing on the next object

22 and on the next sequentially.

1   In all of the above cases, the automatic zoom can more naturally choose

2   to home-in on the person with the brightest emission or reflection, and then proceed to

3   the next brightness and so forth. This would be a form of an intensity/time selection

4   multiplex zoom system.

5   The relative positioning of the input camera with respect to the area under

6   surveillance will effect the accuracy by which the image raster analyzer segments each

7   image. In this preferred embodiment, it is beneficial for the input camera to view the

8   area under surveillance from a point located directly above, e.g., with the input camera

9   mounted high on a wall, a utility tower, or a traffic light support tower. The height of

10   the input camera is preferably sufficient to minimize occlusion between the input camera

11   and the movement of the individuals under surveillance.

12   Once the objects within each sampled video frame are segmented (i.e.,

13   detected and isolated), an analysis is made of the detailed movements of each object

14   located within each particular segment of each image, and their relative movements with

15   respect to the other objects.

16   Each image frame segment, once digitized, is stored in a frame by frame

17   memory storage of picture processing means 12.   Each frame from the picture input

18   means 10 is subtracted from a previous frame already stored in processing means 12

19   using any conventional differencing process. The differencing process involving multiple

20   differencing steps takes place in the processing section 12. The resulting difference

21   signal (outputted from the differencing sub-section of means 12) of each image indicates

22   all the changes that have occurred from one frame to the next. ~~(See Reference 2 above,~~

B.

1   (at page 6.) These changes include any movements of the individuals located within the

2   segment and any movements of their limbs, e.g., arms.

3   Referring to Fig.3, a collection of differencing signals for each moved

4   object of subsequent sampled frames of images (called a "track") allows a determination

5   of the type, speed and direction (vector) of each motion involved, processing which will

6   extract acceleration, i.e., note of change of velocity: and change in acceleration with

7   respect to time (called "jerkiness"), and correlating this with stored signatures of known

8   physical criminal acts. For example, subsequent differencing signals may reveal that an

9   individual's arm is moving to a high position, such as the upper limit of that arm's

10   motion, i.e., above his head) at a fast speed. This particular movement could be

11   perceived, as described below, as a hostile movement with a possible criminal activity

12   requiring the expert analysis of security personnel.

13   The intersection of two tracks indicates the intersection of two moved

14   objects. The intersecting objects, in this case, could be merely the two hands of two

15   people greeting each other, or depending on other characteristics, as described below,

16   the intersecting objects could be interpreted as a fist of an assailant contacting the face

17   of a victim in a less friendly greeting. In any event, the intersection of two tracks

18   immediately requires further analysis and/or the summoning of security personnel. But

19   the generation of an alarm, light and sound devices located, for example, on a monitor

20   will turn a guard's attention only to that monitor, hence the labor savings. In general

21   however, friendly interactions between individuals is a much slower physical process than

22   is a physical assault vis-a-vis body parts of the individuals involved. Hence, friendly

1 interactions may be easily distinguished from hostile physical acts using current low pass

2 and high pass filters, and current pattern recognition techniques based on experimental

3 reference data.

4 When a large number of sensors (called a sensor suite) are distributed over

5 a large number of facilities, for example, a number of ATMs (automatic teller machines),

6 associated with particular bank branches and in a particular state or states and all

7 operated under a single bank network control, then only one monitor is required.

8 A commercially available software tool may enhance object-movement

9 analysis between frames (called optical flow computation). ~~(See Ref. 3 and 4 above at~~

10 ~~page 6)~~ With optical flow computation, specific (usually bright) reflective elements,

11 called farkles, emitted from the clothing and/or the body parts of an individual of one

12 frame are subtracted from a previous frame. The bright portions will inherently provide

13 sharper detail and therefore will yield more accurate data regarding the velocities of the

14 relative moving objects. Additional computation, as described below, will provide data

15 regarding the acceleration and even change in acceleration or "jerkiness" of each moving

16 part sampled.

17 The physical motions of the individuals involved in an interaction, will be

18 detected by first determining the edges of the of each person imaged. And the

19 movements of the body parts will then be observed by noting the movements of the edges

20 of the body parts of the individuals involved in the interaction. The differencing process

21 will enable the determination of the velocity and acceleration and rate of acceleration of

22 those body parts.

1   The now processed signal is sent to comparison means 14 which compares

2   selected frames of the video signals from the picture input means 10 with "signature"

3   video signals stored in memory 16. The signature signals are representative of various

4   positions and movements of the body ports of an individual having various levels of

5   criminal intent. The method for obtaining the data base of these signature video signals

6   in accordance with another aspect of the invention is described in greater detail below.

7   If a comparison is made positive with one or more of the signature video

8   signals, an output "alert" signal is sent from the comparison means 14 to a controller 18.

9   The controller 18 controls the operation of a secondary, high resolution picture input

10   means (video camera) 20 and a conventional monitor 22 and video recorder 24. The

11   field of view of the secondary camera 20 is preferably at most, the same as the field of

12   view of the primary camera 10, surveying a second observation area. The recorder 24

13   may be located at the site and/or at both a law enforcement facility (not shown) and

14   simultaneously at a court office or legal facility to prevent loss of incriminating

15   information due to tampering.

16   The purpose of the secondary camera 20 is to provide a detailed video

17   signal of the individual having assumed criminal intent and also to improve false positive

18   and false negative performance. This information is recorded by the video recorder 24

19   and displayed on a monitor 22. An alarm bell or light (not shown) or both may be

20   provided and activated by an output signal from the controller 20 to summon a supervisor

21   to immediately view the pertinent video images showing the apparent crime in progress

22   and access its accuracy.

1   In still another embodiment of the invention, a VCR 26 is operating

2   continuously (using a 6 hour loop-tape, for example). The VCR 26 is being controlled

3   by the VCR controller 28. All the "real-time" images directly from the picture input

4   means 10 are immediately recorded and stored for at least 6 hours, for example. Should

5   it be determined that a crime is in progress, a signal from the controller 18 is sent to the

6   VCR controller 28 changing the mode of recording from tape looping mode to non-

7   looping mode. Once the VCR 26 is changed to a non-looping mode, the tape will not

8   re-loop and will therefore retain the perhaps vital recorded video information of the

9   surveyed site, including the crime itself, and the events leading up to the crime.

10   When the non-looping mode is initiated, the video signal may also be

11   transmitted to a VCR located elsewhere; for example, at a law enforcement facility and,

12   simultaneously to other secure locations of the Court and its associated offices.

13   Prior to the video signals being compared with the "signature" signals

14   stored in memory, each sampled frame of video is "segmented" into parts relating to the

15   objects detected therein. To segment a video signal, the video signal derived from the

16   vidicon or CCD/TV camera is analyzed by an image raster analyzer. Although this

17   process causes slight signal delays, it is accomplished nearly in real time.

18   At certain sites, or in certain situations, a high resolution camera may not

19   be required or otherwise used. For example, the resolution provided by a relatively

20   simple and low cost camera may be sufficient. Depending on the level of security for

21   the particular location being surveyed, and the time of day, the length of frame intervals

22   between analyzed frames may vary. For example, in a high risk area, every frame from

14

1   the CCD/TV camera may be analyzed continuously to ensure that the maximum amount

2   of information is recorded prior to and during a crime. In a low risk area, it may be

3   preferred to sample perhaps every 10 frames from each camera, sequentially.

4   If, during such a sampling, it is determined that an abnormal or suspicious

5   event is occurring, such as two people moving very close to each other, then the system

6   would activate an alert mode wherein the system becomes "concerned and curious" in

7   the suspicious actions and the sampling rate is increased to perhaps every 5 frames or

8   even every frame. As described in greater detail below, depending on the type of system

9   employed (i.e., video only, audio only or both), during such an alert mode, the entire

10   system may be activated wherein both audio and video system begin to sample the

11   environment for sufficient information to determine the intent of the actions.

12   Referring to Fig. 2, several frames of a particular camera output are

13   shown to illustrate the segmentation process performed in accordance with the invention.

14   The system begins to sample at frame K and determines that there are four objects

15   (previously determined to be people, as described below), A-D located within a particular

16   zone being policed. Since nothing unusual is determined from the initial analysis, the

17   system does not warrant an "alert" status. People A, B, and D are moving according to

18   normal, non-criminal intent, as could be observed.

19   A crime likelihood is indicated when frames K+10 through K+13 are

20   analyzed by the differencing process. And if the movement of the body parts indicate

21   velocity, acceleration and "jerkiness" that compare positively with the stored digital

1   signals depicting movements of known criminal physical assaults, it is likely that a crime

2   is in progress here.

3   Additionally, if a high velocity of departure is indicated when person C

4   moves away from person B, as indicated in frames K+15 through K+17, a larger level

5   of confidence, is attained in deciding that a physical criminal act has taken place or is

6   about to.

7   An alarm is generated the instant any of the above conditions is

8   established. This alarm condition will result in sending in Police or Guards to the crime

9   site, activating the high resolution CCD/TV camera to record the face of the person

10 committing the assault, a loud speaker being activated automatically, playing a recorded

11 announcement warning the perpetrator the seriousness of his actions now being

12 undertaken and demanding that he cease the criminal act. After dark a strong light will

13 be turned on automatically. The automated responses will be actuated the instant an

14 alarm condition is determined by the processor. Furthermore, an alarm signal is sent to

15 the police station, and the same video signal of the event is transmitted to a court

16 appointed data collection office, to the Public Defender's office and the District

17 Attorney's Office.

18   As described above, it is necessary to compare the resulting signature of

19 physical body parts motion involved in a physical criminal act, that is expressed by

20 specific motion characteristics (i.e., velocity, acceleration, change of acceleration), with

21 a set of signature files of physical criminal acts, in which body parts motion are equally

1   involved. This <u>comparison,</u> is commonly referred to as pattern matching and is part of

2   the pattern recognition process.

3   Files of physical criminal acts, which involve body parts movements such

4   as hands, arms, elbows, shoulder, head, torso, legs, and feet, can be reviewed to

5   ascertain this pattern. In addition, a priority can be set by experiments and simulations

6   of physical criminal acts gathered from "dramas" that are enacted by professional actors,

7   the data gathered from experienced muggers who have been caught by the police as well

8   as victims who have reported details of their experiences will help the actors perform

9   accurately. Video of their motions involved in these simulated acts can be stored in

10   digitized form and files prepared for signature motion of each of the body parts involved,

11   in the simulated physical criminal acts.

12   In another embodiment, the above described Abnormality Detection System

13   includes an RF-ID (Radio Frequency Identification) tag or card to assist in the detection

14   and tracking of individuals within the field of view of a camera. Such cards or tags

15   could be used by authorized individuals to respond when queried by the RF interrogator.

16   The response signal of the tags propagation pattern which is adequately registered with

17   the video sensor. The card or tag, when sensed in video, would be assumed friendly and

18   authorized. This information would simplify the segmentation process.

19   A light connected to each RF-ID card will be turned ON, when a positive

20   response to an interrogation signal is established. The light will appear on the computer

21   generated grid (also on the screen of the monitor) and the intersection of tracks clearly

22   indicated, followed by their physical interaction. But also noted will be the intersection

between the tagged and the untagged individuals. In all of such cases, the segmentation process will be simpler.

There are many manufacturers of RF-ID cards and Interrogators, three major ones are, The David Sarnoff Research Center of Princeton, New Jersey, AMTECH of Dallas, Texas and MICRON Technology of Boise, Idaho.

The applications of the present invention include banks, ATMs, hotels, schools, residence halls and dormitories, office and residential buildings, hospitals, sidewalks, street crossings, parks, containers and container loading areas, shipping piers, train stations, truck loading stations, airport passenger and freight facilities, bus stations, subway stations, theaters, concert halls, sport arenas, libraries, churches, museums, stores, shopping malls, restaurants, convenience stores, bars, coffee shops, gasoline stations, highway rest stops, tunnels, bridges, gateways, sections of highways, toll booths, warehouses, and depots, factories and assembly rooms, law enforcement facilities including jails. Any location or facility, civilian or military, requiring security would be a likely application.

Further applications of this invention are in moving platforms: automobiles, trucks, buses, subway cars,train cars, both freight and passenger, boats, ships (passenger and freight), tankers, service and construction vehicles, on and off-road, containers and their carriers, and airplanes, and also in equivalent military and sensitive mobile platforms.

As a deterrence to car-jacking a tiny CCD/TV camera hidden in the ceiling or the rearview mirror of the car, and focussed through a pin hole lens to the driver's

1   seat, may be connected to the video processor to record the face of the drive. The

2   camera is triggered by the automatic word recognition processor that will identify the

3   well known expressions commonly used by the car-jacker. The video picture will be

4   recorded and then transmitted via cellular phone in the car. Without a phone, the short

5   video recording of the face of the car-jacker will be held until the car is found by the

6   police, but now with the evidence (the picture of the car-jacker) in hand.

7   In this present surveillance system, the security personnel manning the

8   monitors are alerted only to video images which show suspicious actions (criminal

9   activities) within a prescribed observation zone. The security personnel are therefore

10   used to access the accuracy of the crime and determine the necessary actions for an

11   appropriate response. By using computers to effectively filter out all normal and

12   noncriminal video signals from observation areas, fewer security personnel are required

13   to survey and "secure" a greater overall area (including a greater number of observation

14   areas, i.e., cameras).

15   It is also contemplated that the present system could be applied to assist

16   blind people "see". A battery operated portable version of the video system would

17   automatically identify known objects in its field of view and a speech synthesizer would

18   "say" the object. For example, "chair", "table", etc. would indicate the presence of a

19   chair and a table.

20   Depending on the area to be policed, it is preferable that at least two and

21   perhaps three cameras (or video sensors) are used simultaneously to cover the area.

22   Should one camera sense a first level of criminal action, the other two could be

1   manipulated to provide a three dimensional perspective coverage of the action. The three

2   dimensional image of a physical interaction in the policed area would allow observation

3   of a greater number of details associated with the steps: accost, threat, assault, response

4   and post response. The conversion process from the two dimensional image to the three

5   dimensional image is achieved by use of the known Radon transform.

6   In the extended operation phase of the invention as more details of the

7   physical variation of movement characteristics of physical threats and assaults against a

8   victim and also the speaker independent (male, female of different ages groups) and

9   dialect independent words and terse sentences, with corresponding responses, will enable

10   automatic recognition of a criminal assault, without he need of guard, unless required by

11   statutes and other external requirements.

12   In another embodiment of the present invention, both video and acoustic

13   information is sampled and analyzed. The acoustic information is sampled and analyzed

14   in a similar manner to the sampling and analyzing of the above-described video

15   information. The audio information is sampled and analyzed in a manner shown in Fig.

16   4, and is based on prior art. ~~(See references 6 and 7 at page 7 above.)~~

17   The employment of the audio speech band, with its associated Automatic

18   Speech Recognition (ASR) system, will not only reduce the false alarm rate resulting

19   from the video analysis, but can also be used to trigger the video and other sensors if the

20   sound threat predates the observed threat.

21   Referring to Fig. 4, a conventional automatic word recognition system is

22   shown, including an input microphone system 40, an analysis subsystem 42, a template

subsystem 44, a pattern comparator 46, and a post-processor and decision logic subsystem 48.

In operation, upon activation, the acoustic/audio policing system will begin sampling all (or a selected portion) of nearby acoustic signals. The acoustic signals will include voices and background noise. The background noise signals are generally known and predictable, and may therefore be easily filtered out using conventional filtering techniques. Among the expected noise signals are unfamiliar speech, automotive related sounds, honking, sirens, the sound of wind and/or rain.

The microphone input system 40 pick-up the acoustic signals and immediately filter out the predictable background noise signals and amplify the remaining recognizable acoustic signals. The filtered acoustic signals are analyzed in the analysis subsystem 42 which processes the signals by means of digital and spectral analysis techniques. The output of the analysis subsystem is compared in the pattern comparater subsystem 46 with selected predetermined words stored in memory in 44. The post processing and decision logic subsystem 48 generates an alarm signal, as described below.

The templates 44 include perhaps about 100 brief and easily recognizable terse expressions, some of which are single words, and are commonly used by those intent on a criminal act. Some examples of commonly used word phrases spoken by a criminal to a victim prior to a mugging, for example, include: "Give me your money", "This is a stick-up", "Give me your wallet and you won't get hurt"...etc. Furthermore, commonly used replies from a typical victim during such a mugging may also be stored

1 as template words, such as "help", and certain sounds such as shrieks, screams and

2 groans, etc.

3   The specific word templates, from which inputed acoustic sounds are

4 compared with, must be chosen carefully, taking into account the particular accents and

5 slang of the language spoken in the region of concern. Hence, a statistical averaging of

6 the spectral content of each word must be used.

7   The output of the word recognition system shown in Fig. 4 is used as a

8 trigger signal to activate a sound recorder, or a camera used elsewhere in the invention,

9 as described below.

10   The preferred microphone used in the microphone input subsystem 40 is

11 a shot-gun microphone, such as those commercially available from the Sennheiser

12 Company of Frankfurt, Germany. These microphone have a super-cardioid propagation

13 pattern. However, the gain of the pattern may be too small for high traffic areas and

14 may therefore require more than one microphone in an array configuration to adequately

15 focus and track in these areas. The propagation pattern of the microphone system

16 enables better focusing on a moving sound source (e.g., a person walking and talking).

17 A conventional directional microphone may also be used in place of a shot-gun type

18 microphone, such as those made by the Sony Corporation of Tokyo, Japan. Such

19 directional microphones will achieve similar gain to the shot-gun type microphones, but

20 with a smaller physical structure.

1   A feedback loop circuit (not specifically shown) originating in the post

2 processing subsystem 48 will direct the microphone system to track a particular dynamic

3 source of sound within the area surveyed by video cameras.

4   An override signal from the video portion of the present invention will

5 activate and direct the microphone system towards the direction of the field of view of

6 the camera. In other words, should the video system detect a potential crime in

7 progress, the video system will control the audio recording system towards the scene of

8 interest. Likewise, should the audio system detect words of an aggressive nature, as

9 described above, the audio system will direct appropriate video cameras to visually cover

10 and record the apparent source of the sound.

11   A number of companies have developed very accurate and efficient,

12 speaker independent word recognition systems based on a hidden Markov model (HMM)

13 in combination with an artificial neural network (ANN). These companies include IBM

14 of Armonk, NY, AT&T Bell Laboratories, Kurtzweil of Cambridge, MA and Lernout

15 and Hauspie of Belgium.

16   Put briefly, the HMM applies probabilistic statistical procedure in

17 recognizing words. In the training steps, an estimate is made of the means and

18 covariance of the probabilistic model of each word, e.g., those words which are

19 considered likely to be uttered in an interaction. The various ways which any given

20 word is pronounced, permits the spectral parameters of the word to be an effective

21 describer of the model. The steps involved in recognizing an input of an unknown word

22 consists of computing the likelihood that the word was generated by each of the models

developed during the training. The word is considered as "recognized" when its model gives the highest score. Finally, since the words are composed of word units, the evaluation of conditional probabilities of one particular unit followed by the same or another word unit is also part of the computation.

The resulting list of potential words is considerably shorter than the entire list of all spoken words of the English language. Therefore, the HMM system employed with the present invention allows both the audio and video systems to operate quickly and use HMM probability statistics to predict future movements or words based on an early recognition of initial movements and word stems.

The HMM system may be equally employed in the video recognition system. For example, if a person's arm quickly moves above his head, the HMM system may determine that there is a high probability that the arm will quickly come down, perhaps indicating a criminal intent.

While certain embodiments of the invention have been described for illustrative purposes, it is to be understood that there may be various other modifications and embodiments within the scope of the invention as defined by the following claims.